Supplementary Methods

Detection of the synthesized ORF products using ReAsH.

Synthesis of the ORF product was investigated by adding 5 μM ReAsH(1) (Invitrogen), a fluorescein derivative with two As (III) substituents that fluoresces only after binding of arsenic to the tetracysteine tag (CCPGCC), to the *in vitro* translation system after the GFP synthesis reaction, and the fluorescence intensity of the reacted ReAsH was measured with an Mx3005P instrument (Stratagene) at 37°C using the appropriate filter set (585/610 nm). Note that ReAsH was first mixed with an equal volume of 1 M dithiothreitol (DTT) and incubated for 15 min at 4°C prior to addition to the *in vitro* translation system. The ReAsH fluorescence signal was detected with more than 98% of ORFs tested (Supplementary Fig. 2A).

The red fluorescence intensity obtained using ReAsH is not a universal measure of the protein concentration, because it differs depending on the surrounding sequence(2). For example, three purified ORF proteins (Tig, ClpP, and ClpX) containing a tetracysteine tag were labeled with ReAsH and the correlation between protein concentration and red fluorescence intensity was investigated. While we observed a linear correlation between protein concentration and red fluorescence intensity for each protein, we observed differences in the correlation (*i.e.*, slope) among the proteins (Supplementary Fig. 2B). Nevertheless, the concentrations of these three proteins synthesized *in vitro* were estimated to be in the μM range from their respective standard curves (Supplementary Fig. 2B) and also from the band intensity observed after SDS-PAGE (data not shown). Therefore, although we did not estimate the concentrations of all ORF product synthesized *in vitro*, a substantial fraction are likely to be in the μM range.

Cloning, expression, and purification of Tig, ClpP, and ClpX proteins.

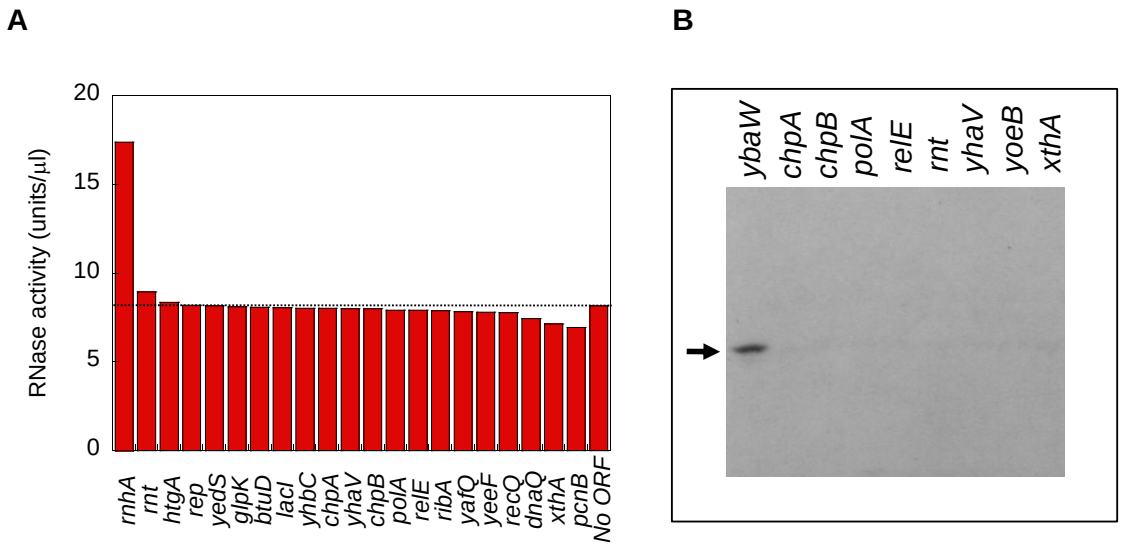
Plasmids of the ASKA library containing *tig*, *clpP*, and *clpX* genes were amplified with the primers pqe2+ and cDNA-lumio-stop2. The PCR conditions were the same as described in the Experimental procedures section. The resulting DNA fragments were cloned into the pCR4-TOPO vector using a TOPO TA cloning kit

(Invitrogen). The plasmids thus obtained were transformed into *E. coli* XL10-gold cells (Stratagene), which were then grown in LB medium at 37°C. On reaching an OD_{600} of 0.8, IPTG was added to a final concentration of 0.1 mM, incubated for a further 3 h, and harvested. His-tagged ORF proteins were purified using immobilized metal ion affinity chromatography (IMAC) as described in the Experimental procedures section.

RNase activity assay

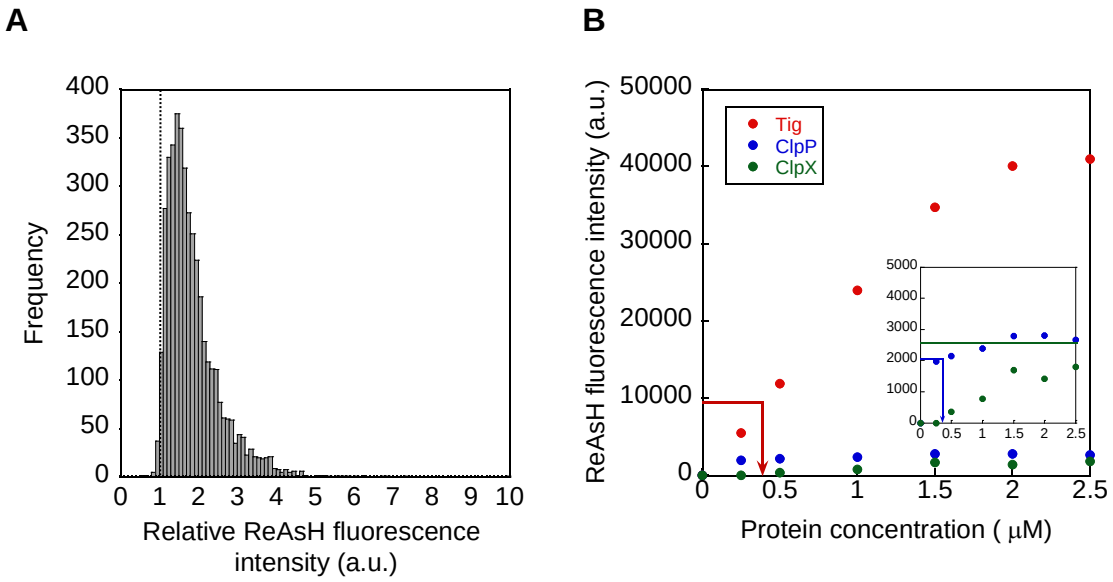
RNase activity was measured using an RNase Alert Kit (Ambion), which gives a fluorescent signal in the presence of RNase activity. ORF products were synthesized using the *in vitro* translation system as described in the Experimental procedures section. Aliquots of the synthesized protein (2 µl) were added to 18 µl of RNase Alert mixture and fluorescence was monitored continuously in Mx3005P at 37°C for 1 h. A 585/610 nm filter set was used. RNase activity was estimated from the standard curve obtained with a dilution series of purified RNaseA (Ambion).

Supplementary Figures



Supplementary Figure 1

Properties of deleterious components. (A) RNase activity assay of deleterious components. RNase activity was measured using an RNase Alert Kit (Ambion). ORF products were synthesized *in vitro* as described in the Experimental procedures section. Aliquots of the reaction mixture were used to measure the RNase activity. RNase activities of the ORF products were estimated using purified RNaseA as a standard. Rnt (RNase T) and RnhA (RNase HI) showed detectable activity, whereas others were similar to the control without any ORF product. (B) Deleterious components suppressed GFP synthesis as revealed by SDS-PAGE followed by western blotting. An arrow indicates the position of GFP. Two independent experiments gave similar results.



Supplementary Figure 2

Expression of 4194 ORF products evaluated by ReAsH. (A) Frequency distribution of the ReAsH fluorescence intensity values. ReAsH fluorescence intensity values were divided by the fluorescence intensity obtained by carrying out the protein synthesis reaction without any ORF genes. The bold line shows a value of 1, and thus the fluorescence intensity when expression was not observed. More than 98% of the ORFs showed values higher than 1. (B) Correlation between ReAsH fluorescence intensity and protein concentration. Here, three proteins chosen arbitrarily, Tig (trigger factor), ClpP (subunit of serine proteases), and ClpX (subunit of serine proteases), carrying the tetra-Cys tag at the C terminus were overexpressed and purified from *E. coli* cells. These were added to the *in vitro* translation system containing ReAsH, and the fluorescence intensity was measured. The lines with arrows indicate the fluorescence intensity obtained for each protein, and the estimated concentrations of synthesized proteins were 0.37 and 0.65 μM with Tig and ClpX, respectively. ClpP gave values greater than 2.5 μM.

Legends to Supplementary Tables

Supplementary Table 1

Gene names were assigned as described by Riley *et al.*(3) based on the genome of *E. coli* K-12. Those that are unique to the W3110 strain were assigned based on the Locustag of W3110. Locustags of each gene on the MG1655 genome (b number), W3110 genome (JW number), and K-12 genome (ECK number) are listed (3).

Types of gene products are listed as in Table 3 of Riley *et al.*(3) e: enzyme; pe: predicted enzyme; t: transporter; pt: predicted transporter; r: regulator; pr: predicted regulator; m: membrane protein; pm: predicted membrane protein; f: factor; pf: predicted factor; s: structural component; ps: predicted structural component; c: carrier protein; pc: predicted carrier protein; lp: lipoprotein; cp: protein involved in cell process; l: leader peptide; su: pseudogenes in common between strains; h: phage/IS in common between strains; d: protein for which only partial information is available. Gene product description was as described by Riley *et al*(3). The term “fused” is used in the description for the fused proteins and their activities. The symbol “-!-” is used to separate the activities of such proteins.

Supplementary Table 2

Comparison of two *E. coli* protein-protein interaction networks(4,5). At present, two sets of data are available for the *E. coli* PPI network (4,5), one of which was obtained in the group of one of the authors (H.M.) Number of proteins, average degree (number of interacting partners per protein) and average clustering coefficient of each protein group in the two networks are shown in the table. Clustering coefficient (6) is defined by $2E_n/d$ (d-1) where d and E_n are the number of interacting proteins (degree) and the number of links between the neighboring proteins, respectively. The average clustering coefficients of the networks detected in Butland *et al.* were double of those reported by Arifuzzaman *et al.*, while the average degree was not markedly different between the two data sets. The lower value of the clustering coefficient indicates

that PPI data of Arifuzzaman *et al.* tend to miss the interactions within large complexes, while these data are suitable for identifying potential binding partners. Furthermore, in the latter study, the bait proteins were overexpressed when acquiring the PPI data. Hence, we used the data of Butland *et al.* consisting of 1255 proteins, which maintained the cellular stoichiometry when acquiring the data and included the interactions within large complexes, for further analysis.

Supplementary Table 3

The number of proteins used for the PPI network analysis.

References

1. Adams, S. R., Campbell, R. E., Gross, L. A., Martin, B. R.*, et al.* (2002) New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications *J. Am. Chem. Soc.* **124**(21), 6063-6076
2. Martin, B. R., Giepmans, B. N., Adams, S. R., and Tsien, R. Y. (2005) Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity *Nat. Biotechnol.* **23**(10), 1308-1314
3. Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K.*, et al.* (2006) Escherichia coli K-12: a cooperatively developed annotation snapshot--2005 *Nucleic Acids Res.* **34**(1), 1-9
4. Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K.*, et al.* (2006) Large-scale identification of protein-protein interaction of Escherichia coli K-12 *Genome Res.* **16**(5), 686-691
5. Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W.*, et al.* (2005) Interaction network containing conserved and essential protein complexes in Escherichia coli *Nature* **433**(7025), 531-537
6. Watts, D. J., and Strogatz, S. H. (1998) Collective dynamics of 'small-world' networks *Nature* **393**(6684), 440-442